PROCEEDINGS OF TDWG

Biodiversity Information Science and Standards 1: e19825
doi: 10.3897/tdwgproceedings.1.19825

OPEN ACCESS

Conference Abstract

Managing data quality in GBIF: status and plans

Christian Gendreau‡, Dmitry Schigel‡

‡ Global Biodiversity Information Facility - Secretariat, Copenhagen, Denmark

Corresponding author: Christian Gendreau (christiangendreau@gmail.com)

Received: 26 Jul 2017 | Published: 27 Jul 2017

Citation: Gendreau C, Schigel D (2017) Managing data quality in GBIF: status and plans. Proceedings of TDWG 1: e19825. https://doi.org/10.3897/tdwgproceedings.1.19825

Abstract

The data quality topic is getting some traction within the biodiversity community. As a data aggregator and one of the primary source for biodiversity data, GBIF needs to adapt, enhance and expand its current data quality activities. For users unfamiliar with nature and characteristics of the globally aggregated data, data quality remains one of the major concerns and the barriers for use. For expert users, data modifications to ensure fitness for use remain a time and effort consuming activity. Data quality and credit should be the primary concerns of data publishers worldwide, but the practices vary.

The presentation will cover the current state of the GBIF parsing and interpretation with a focus on the current data quality flags that are applied, and on how to make use of data quality flags. In addition, the recent development on the GBIF Data Validator enables a dataset to be parsed and interpreted before its publication online. By using Data Validator, some errors and enhancement possibilities can be detected and, possibly, fixed before the publication of the data through GBIF.org.

Major effort has been put into data quality documentation and solutions over the last years among different aggregators, institutions and contributors within the community. Based on the contributions from TDWG Biodiversity Data Quality (BDQ) Interest Group, especially from Task Group 2: Data Quality Tests and Assertions as well as the implementations available in the biodiversity informatics community, we will present an initial plan to merge different informatic code bases together in order to make it available to the community into a more lightweight form.

© Gendreau C, Schigel D. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Keywords

GBIF

Presenting author

Christian Gendreau